Optimizing Zomato Delivery Time through Time Prediction Machine Learning

Self Introduction

Zomato

Education

Universitas Islam Indonesia (2021-2025)
Bachelor in Chemical Engineering
- Designed Preliminary Design of Polyethylene Terephthalate from Ethylene Glycol and Terephthalic Acid

dibimbing.id (2025-Present)
Data Science & AI Machine Learning
- Learning data science methodology, python, SQL, and data visualization.

Work Experience

Laboratory Assistant
Universitas Islam Indonesia
- Delivered 8 modules of practical sessions on Basic Chemical Engineering Laboratory

Process Engineer Internship
PT. Pertamina VI Balongan
- Evaluating splitter column of propylene recovery unit based on real data

Mita Dian Cahyani

Project Overview

Zomato

A/B Testing for E-commerce: Optimizing Speaker Sales Funnels

- **Situation**: A gap in website landing page that result in different revenue
- **Goal**: Efficiently tailor website to drive revenue
- **Action**: Apply website according to customer's preferences
- **Impact**: Increase revenue in 54.4% and conversion rate by 31.5%

Looker e-Commerce Revenue Growth

- **Situation:** Website traffic is declining
- **Goal:** Identify causes of traffic drop
- **Action:** Optimize AdWords on Firefox in key regions (KY, Pais Vasco, WA, Scotland)
- **Impact:** Targeted traffic growth with efficient ad spend

Airline Customer Segmentation (Churn Reduction)

- **Situation**: 46.9% of customers are at risk of churn
- **Goal**: Reduce churn through targeted segmentation
- **Action**: K-means clustering to analyze customer behavior
- **Impact**: 10% recovery of at-risk customers saves $2M

Table of Contents

Zomato

Project Strategy & Objectives

Zomato

Customer Segmentation

Fast and reliable delivery is a key driver of customer satisfaction in food delivery services. Delays reduce trust, increase complaints, and lower the likelihood of repeat orders.

Overview

Zomato operates across metropolitan, urban, and semi-urban areas with varying traffic, distance, and weather conditions.

Goal

To optimize delivery time by using machine learning to predict delivery duration and design strategies that minimize delays.

Objectives

1) **Delivery Time Prediction:** Develop a machine learning model to predict delivery time in real-time

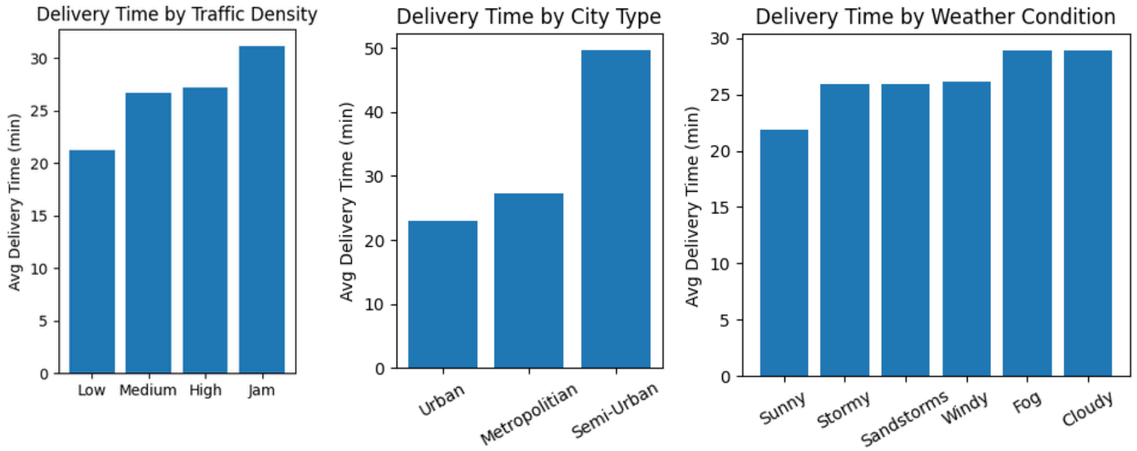
2) **Delivery Time Analysis:** Analyse key factors that impact delivery duration.

3) **Actionable Insights:** Create data-driven operational strategies to reduce delivery time.

Business Problem

Zomato

Problem Statement



Delivery time is affected by external factors which may result inconsistent customer experience. This accurate **delivery time prediction is needed to manage orders**

Why It Matters?

Customer Satisfaction: Improves customer satisfaction and retention by setting accurate delivery expectations

Efficiency: Reduces operational inefficiencies and late deliveries

Stakeholders Benefits

Customers
More accurate delivery time estimates and improved service reliability

Delivery Partners
Better workload distribution and optimized routes

Business Operations
Lower churn, improved efficiency, and higher order completion rates

Data Understanding & Preprocessing

Zomato

This dataset details delivery operations by tracking driver profiles, environmental factors like weather and traffic, and geospatial coordinates. It consists of 45,584 rows and 20 columns.

Analysis link: Google Colab

Tools: Google colab

Key Features

Category	Column Names
Temporal	Order_Date, Time_Ordered, Time_Order_picked, Festival
Geospatial	Restaurant & Delivery Lat/Long, City
Environmental	Weather_conditions, Road_traffic_density
Operational	Vehicle_condition, Multiple_deliveries, Type_of_vehicle

Data Preprocessing

Missing Value
- Numeric column → median
- Categorical column → mode

Change Datatype
- Change ‘order_datetime’ and ‘picked_datetime’ to datetime

Feature Engineering
- ‘**prep_time_min**’ is the time needed for restaurant to prepare order (‘picked_datetime’ **-** ‘order_datetime’
- ‘**delivery_dist_km**’ is the distance between customer and restaurant (using haversine from restaurant lat/long and delivery lat/long)
- ‘**order_hour**’, ‘**order_min**’ abstracted from ‘order_datetime’

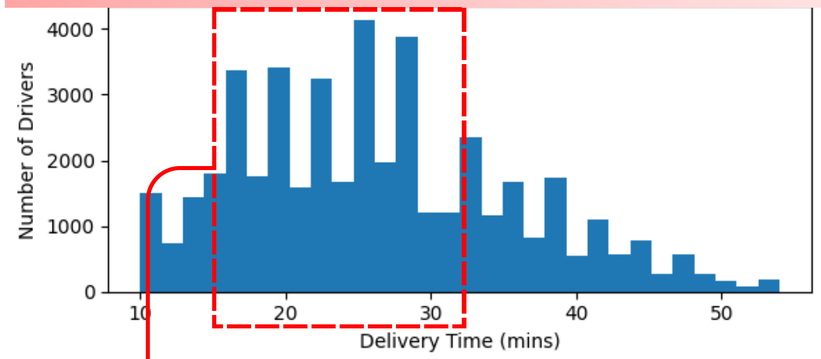
Missing value percentage

delivery_location_longitude	7.99
delivery_location_latitude	7.99
restaurant_longitude	7.99
restaurant_latitude	7.99
delivery_person_ratings	4.19
delivery_person_age	4.07
time_orderd	3.80
city	2.63
multiple_deliveries	2.18
weather_conditions	1.35
road_traffic_density	1.32
festival	0.50

Exploratory Data Analysis

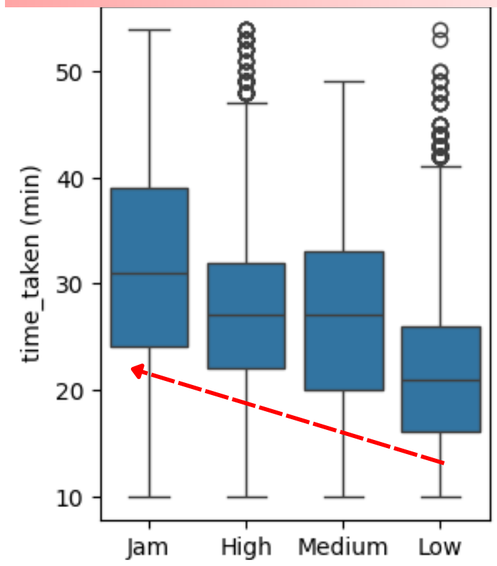
Zomato

Delivery time (min) distribution



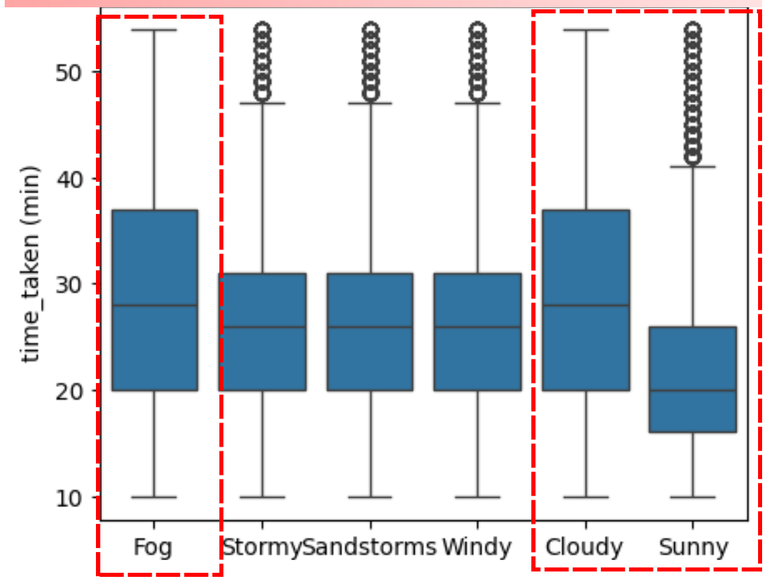
It generally takes 15 to 32 minutes to deliver

Delivery time (min) across road traffic



Busier traffic slows dowm delivery

Delivery time (min) across weather conditions

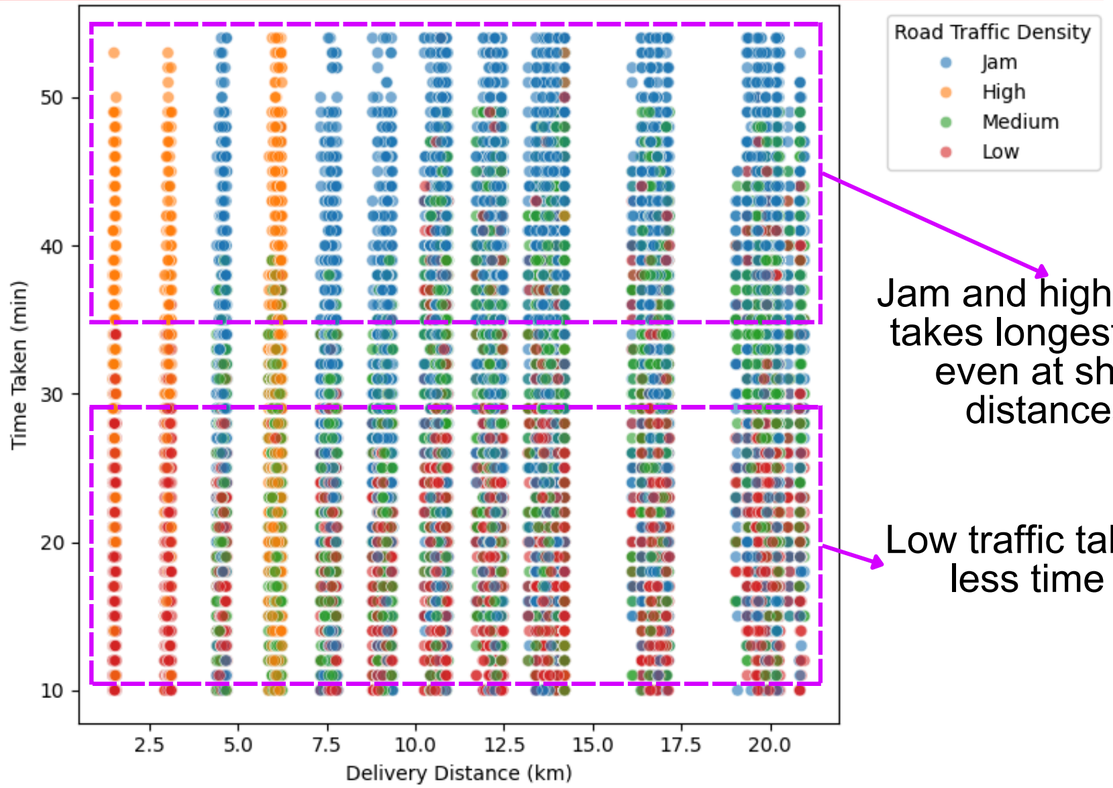


Foggy and cloudy takes longer to deliver, while sunny day is faster

Exploratory Data Analysis

Zomato

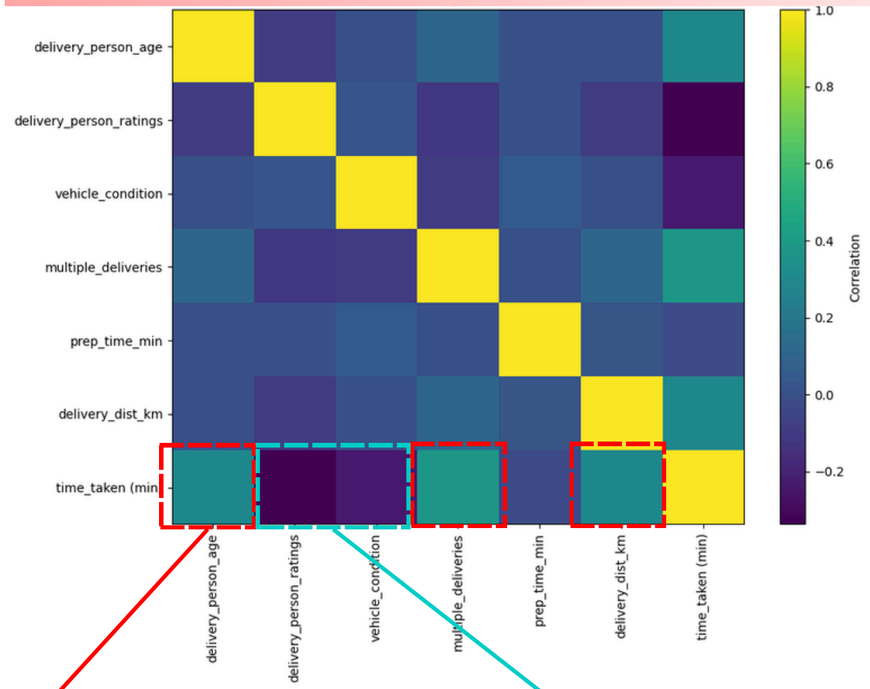
Delivery distance vs time by traffic density



Jam and high traffic takes longest time even at short distances

Low traffic takes less time

Heatmap



Delivery time increase when:

- delivery man is older
- have multiple deliveries
- farther delivery distance

Delivery time decrease when:

- Delivery men have lower ratings
- Vehicle condition is not in best condition

Modelling

Zomato

Model Selection

model	train_MAE	train_RMSE	train_R2	train_MAPE	test_MAE	test_RMSE	test_R2	test_MAPE
RandomForest	2.293537	2.945052	0.901528	10.017312	3.171972	4.028756	0.815543	13.700667
GradientBoosting	3.444700	4.325806	0.787548	15.275560	3.493117	4.376636	0.782313	15.407512
KNN	4.305816	5.463648	0.661083	18.798570	4.751644	6.003080	0.590456	20.739410
Lasso	4.825715	6.070651	0.581594	21.399927	4.823520	6.051273	0.583854	21.344262
Ridge	4.825649	6.070540	0.581609	21.398752	4.823952	6.052136	0.583735	21.344862
LinearRegression	4.825623	6.070539	0.581609	21.398402	4.823969	6.052228	0.583723	21.344622

Why **Random Forest** is chosen?

Main metrics: MAE

- Not overfitting
- Good at handling noise
- Low MAE

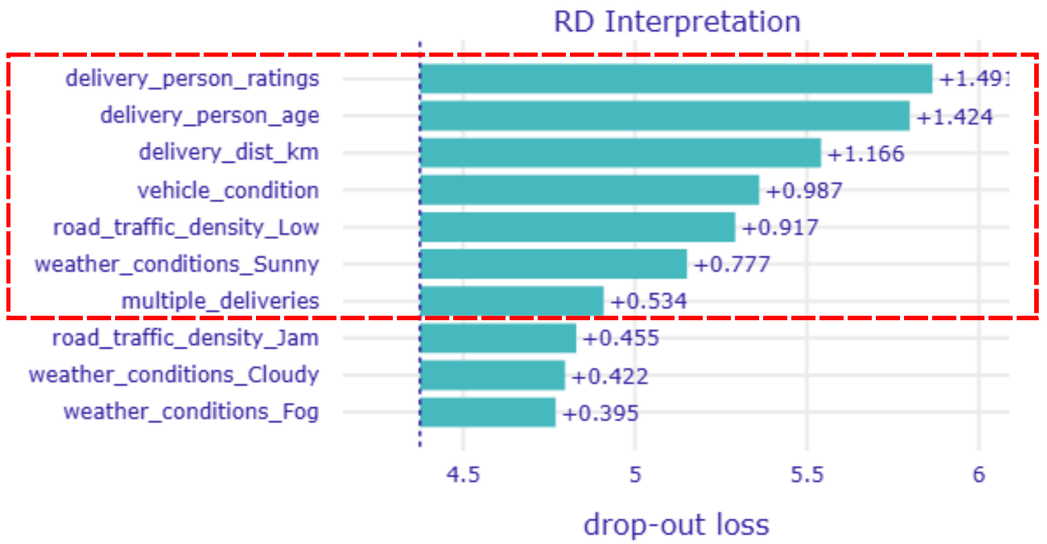
Hyperparameter Tuning

Random Forest Parameter

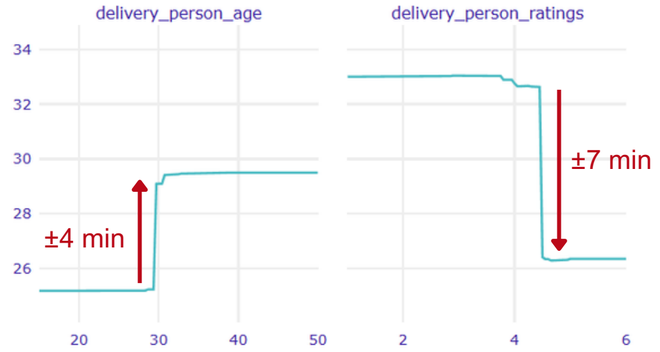
- max_depth=15
- max_features='sqrt'
- min_samples_leaf=30
- n_estimators=150
- n_jobs=-1
- random_state=42

Variable Importance

Interpretation with dalex

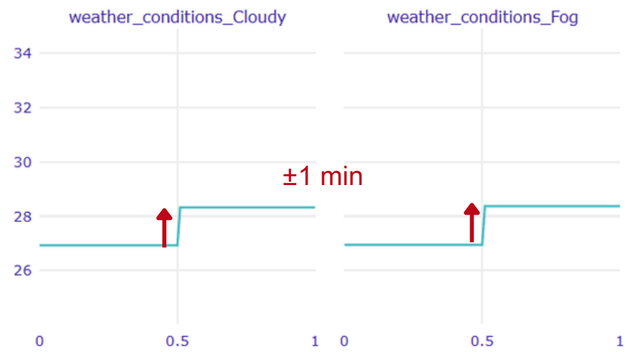


Modelling

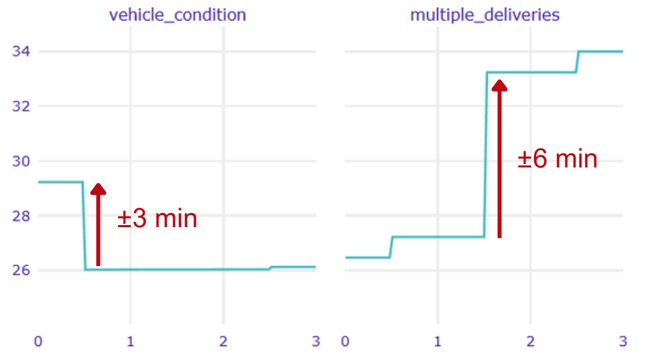


Drivers over 28 take ±4 minutes longer on average

Drivers with ratings above 4 deliver ±7 minutes faster

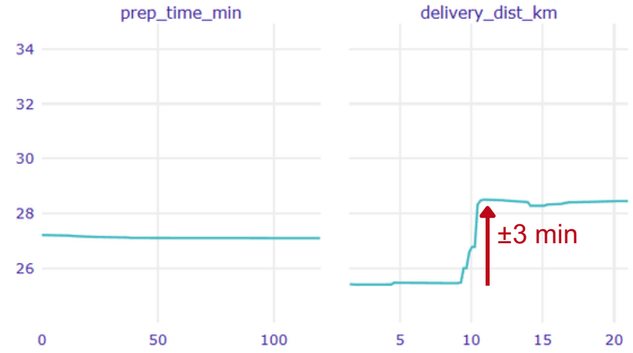


Cloudy and foggy weather ±1 minute longer on average



Better vehicle condition deliver ±3 minutes faster

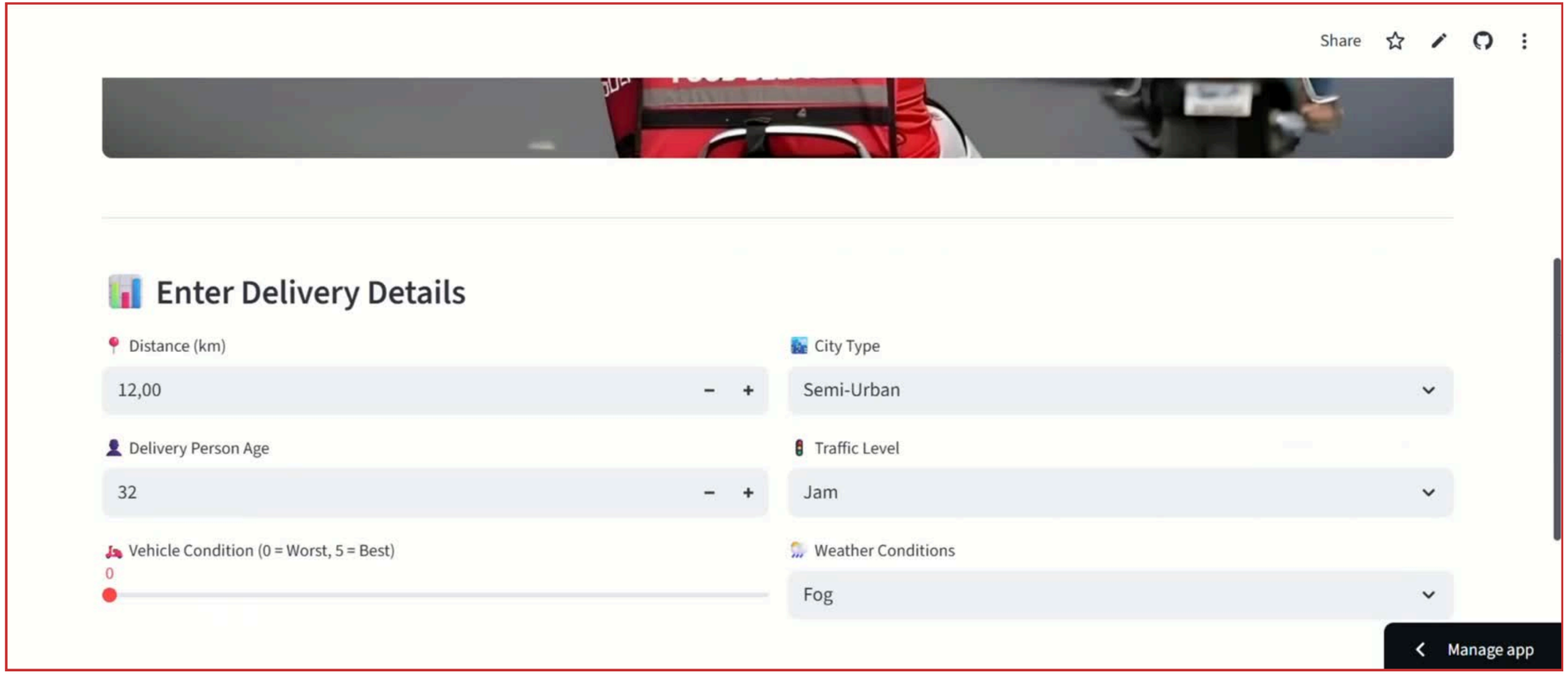
Once drivers have 2 delivery it takes ±6 minutes longer



Delivery distance above 10 km takes ±3 minutes longer

App Deployment

Zomato



Link:

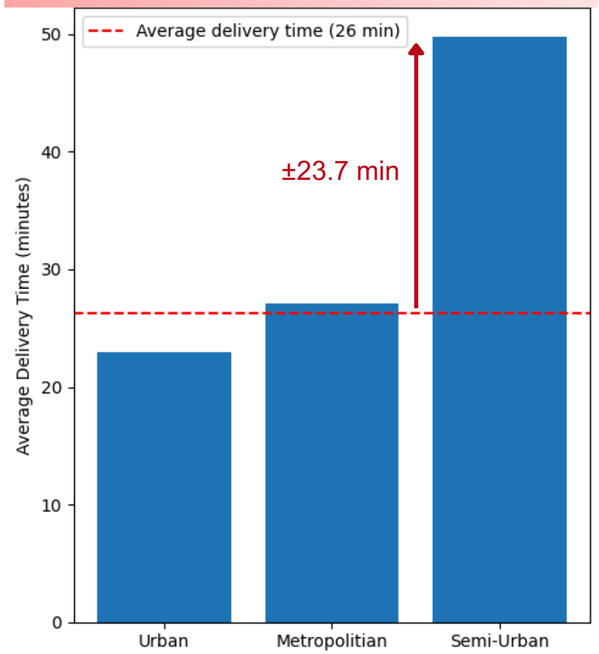
- https://zomatodeliveryprediction-ycaxu2ay76gam83efmn3yv.streamlit.app/
- https://github.com/mitadian/zomato_delivery_prediction

Visual Studio Code Streamlit GitHub

Data Analysis & Insights

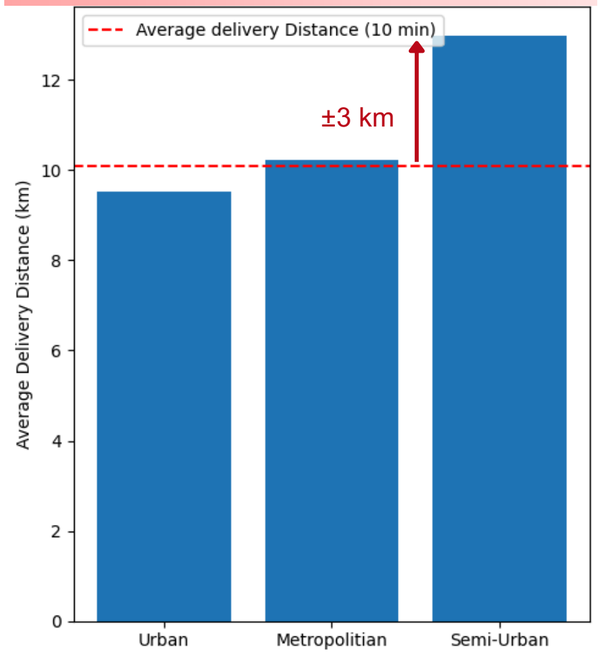
Zomato

(Avg) delivery time by city



Delivery in semi-urban place takes the most longer than average

(Avg) delivery distance by city



Semi-urban delivery is 3km farther than general

Traffic per city (% by row)

road_traffic_density	High	Jam	Low	Medium
city				
Metropolitian	9.859276	32.340827	33.396259	24.403638
Semi-Urban	10.493827	82.716049	0.617284	6.172840
Urban	9.312992	25.974673	41.360056	23.352278

Semi-urban is jammed 82% of time

Weather per city (% by row)

weather_conditions	Cloudy	Fog	Sandstorms	Stormy	Sunny	Windy
city						
Metropolitian	16.881185	18.319890	16.455008	16.660946	15.153595	16.529375
Semi-Urban	28.395062	20.370370	12.345679	12.345679	12.345679	14.197531
Urban	15.295643	16.272809	16.631768	16.920929	19.124539	15.754312

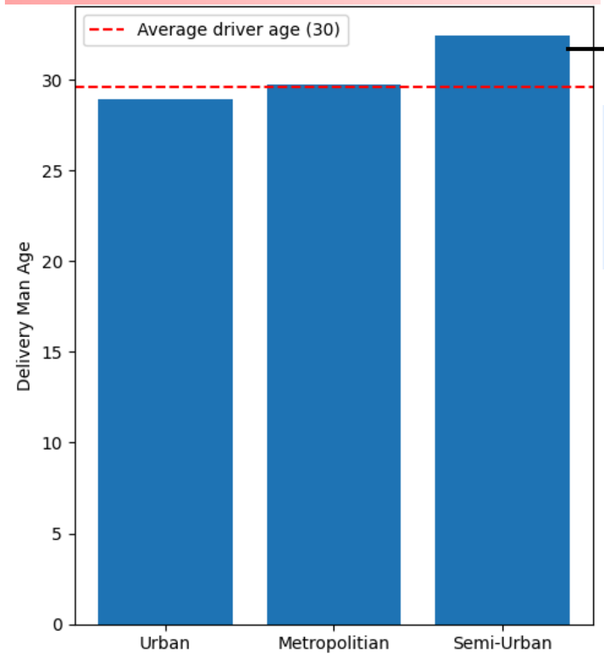
Semi-urban is mostly cloudy and foggy

Semi-urban city is the slowest in delivery which is due to the external factors of far **distance** between restaurant and customer, the **traffic and weather** of the city

Data Analysis & Insights

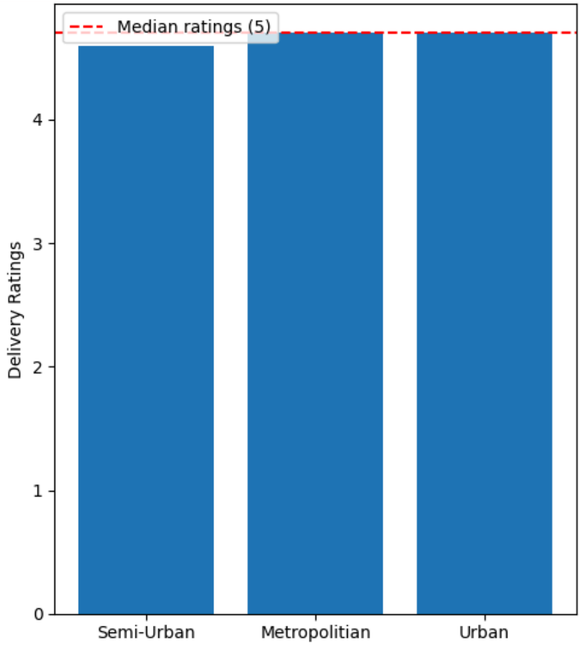
Zomato

(Avg) driver age by city



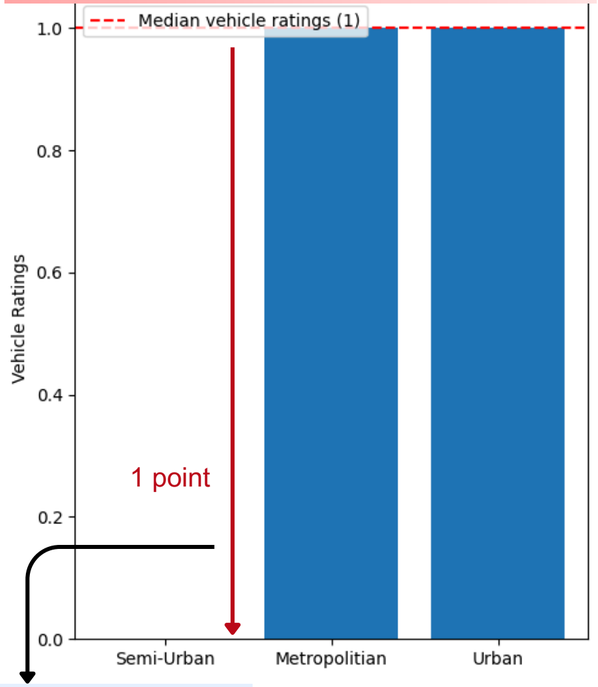
Semi-urban driver is generally 3 years older

(Median) driver ratings by city



The driver ratings across the city is generally in good score

(Median) vehicle ratings by city



Semi-urban driver vehicle condition is not in best condition

Semi-urban city is the slowest in delivery which is due to the internal factors, that is generally **older drivers** and **bad vehicle conditions**

Recommendations and Impact

Zomato

Situation

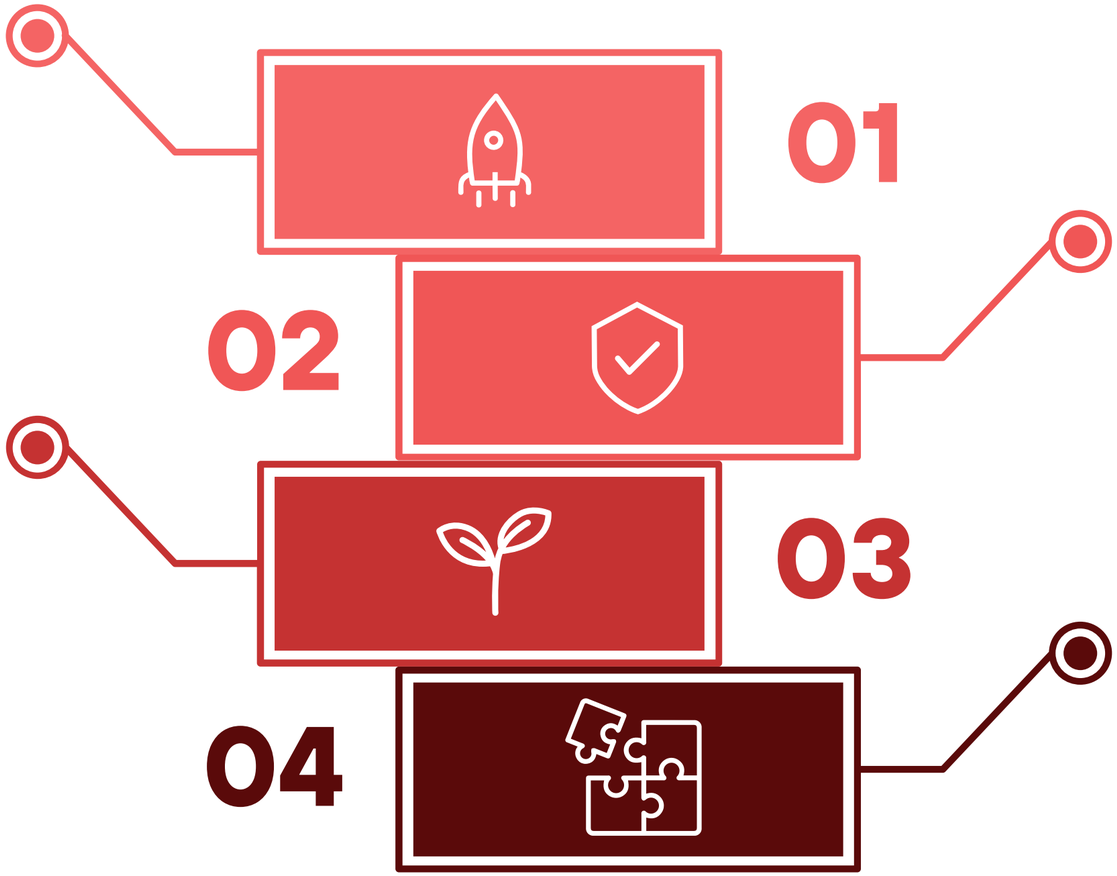
- Semi Urban city are slower in delivery due to delivery distance, weather, traffic, age and vehicle condition

Recommendations

- Allocate productive drivers in Semi Urban city to handle difficult deliveries
- Apply real time delivery prediction in app to maintain customer trust

note:

- Productive drivers have minimum ratings of 5, maximum age of 30 years and minimum vehicle condition of 1
- Difficult deliveries is delivery distance minimum 10 km, jammed road, cloudy and foggy weahter



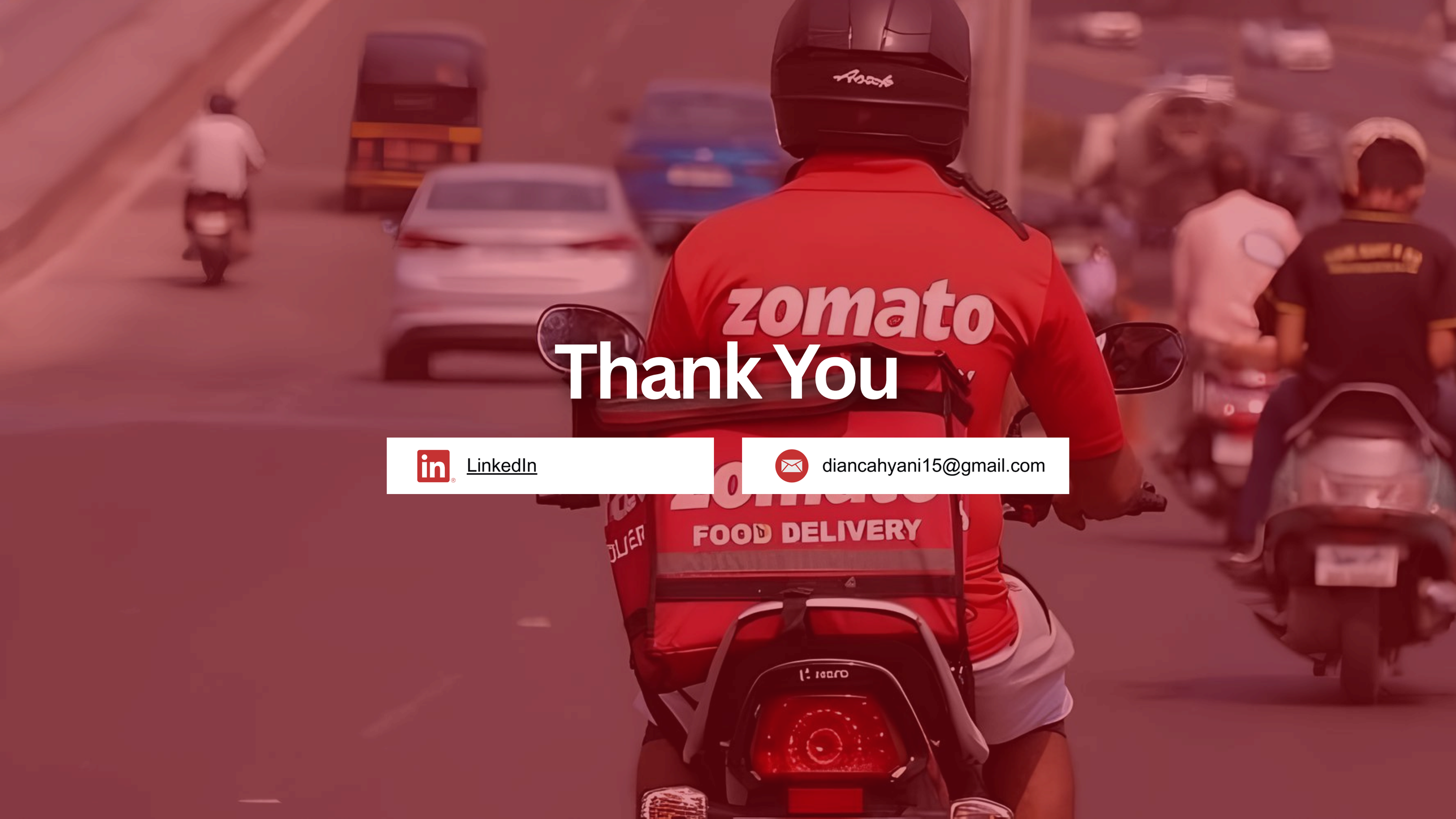
Problem

- Slow down delivery time affecting customer satisfaction

Impact

- Reallocating productive drivers can deliver 3 minutes faster for difficult delivery
- Faster delivery times significantly increase customer satisfaction and loyalty (Natasya Rafmadini et al., 2025)

Natasya Rafmadini, Wahyuningsih Santosa, & Triwulandari Satitidjati Dewayana. (2025). PENGARUH DELIVERY TIME, SERVICE QUALITY, PRODUCT QUALITY AND CONDITION TERHADAP CUSTOMER LOYALTY PADA PELANGGAN GEN Z ECOMMERCE DI JAKARTA MELALUI CUSTOMER SATISFACTION. Jurnal Manajemen Terapan Dan Keuangan, 14(03), 1169–1182. https://doi.org/10.22437/jmk.v14i03.47120



Thank You

LinkedIn

diancahyani15@gmail.com